AMERICAN JOURNAL OF
Public Health
Editorials
April 1987 Established 1911
Volume 77, Number 4

Appraising Health Risk Appraisal

The concept of health risk appraisal (HRA) is generally credited to Dr. Lewis C. Robbins, whose work on cervical cancer and heart disease prevention during the late 1940s led him to the idea that a physician might record a patient's health hazards as a guide to preventive efforts and then to the creation of a simple "health hazard chart" that could give the medical examination a more prospective orientation.[1] A decade later, as Chief of the Cancer Control Program at the Public Health Service Division of Chronic Disease, Robbins directed the preparation of tables of 10-year mortality risk and helped to establish several small demonstration projects in which HRA was used as a medical teaching and practice model.[1,2] By the end of the 1960s, with the application of life insurance actuarial principles to risk assessment and the quantitation of risk multipliers for patient characteristics that affect mortality risk, all the necessary components for quantitative risk appraisal had been created.[3] The 1970 publication of the Robbins and Hall manual *How to Practice Prospective Medicine*,[4] written for the practicing physician, provided a complete HRA package, including questionnaire, risk computations, and feedback strategy.

Although the medical profession largely ignored HRA, the continuing activities of its adherents, the potential for computerization of the risk estimation procedure, commercial interest, and the substantial involvement of Canadian and United States government agencies in the years following the Lalonde Report[5] led to a proliferation of HRA programs[1] and instruments (52 in a recent directory).[6] As many as 5 to 15 million Americans in worksites, universities, community wellness programs, health fairs, and health care organizations may have had an HRA.* About 7 per cent of private-sector worksites with 50 or more employees, and probably a higher percentage of larger worksites, have used HRA.[7]

The accuracy of HRA risk estimates has been a longstanding concern of the technique's developers and of the Society of Prospective Medicine, the professional organization most closely linked to HRA. As the technique gained prominence and attracted the interest of governmental health promotion agencies, systematic reviews of HRA methodology were undertaken.[1,8–10] These and other critiques raised numerous questions about validity of the databases and procedures used in HRA risk estimation. But few empirical evaluations of the adequacy of the HRA risk assessment procedures have been reported.

The study by Smith, *et al*,[11] in this issue of the Journal joins two earlier empirical studies[12,13] in helping to define the boundaries of uncertainty about HRA's validity. In the first such study, Wiley[12] retrospectively computed HRA risk estimates using 13 risk characteristics that had been measured on the Alameda County cohort. HRA differentiated high-, middle-, and low-risk subjects, although it overestimated by 26 deaths per 1,000 the actual mortality experience. As assessed by comparison of log likelihood statistics, HRA's performance nearly matched that of a multiple logistic model.

In the second study, Chaves, *et al*,[13] at the American Institutes for Research found heart disease mortality risk estimates from seven basically similar HRA instruments to be highly (above 0.87) correlated with one another. Assessment of the validity of HRA absolute risk scores, by comparison to an accepted standard, was identified as an appropriate next step. Such an assessment is the objective of the Smith, *et al*,[11] study by the same research group.

*David G. Moriarty, personal communication, December 31, 1986.

A number of difficulties arise in testing the predictive validity of a tool like HRA.[14–16] The most basic problem is that there is no entirely satisfactory validation standard to use—available cohorts are too small, include too few of the prognostic characteristics used in HRAs, and give us information only about past, not present, mortality rates.[14] Indeed, the very experience of participation in a cohort study may change people's mortality experience, so that a true test of HRA predictions may be impossible due to a kind of "uncertainty principle" in which the measurement process changes the mortality experience that it seeks to measure.[15,16]

Precise prediction of disease or mortality by any means is a currently unattainable goal, for such reasons as our incomplete knowledge of the total set of risk factors, their time-dose levels, and the true functional form of their contribution to risk.[14,17] Framingham and similar risk models are generally successful in differentiating high-, medium-, and low-risk individuals and in estimating relative risk, but are much less successful in estimating absolute risk in individuals or across populations.[14] In contrast, measurements applied to individuals should attain higher levels of accuracy than measurements used only in correlational studies, where there is opportunity for random errors to offset one another.[18]

Smith, *et al,*[11] have sidestepped these intractable issues by treating the problem as one of measurement validity rather than predictive validity, and taking as a validation standard coronary heart disease (CHD) mortality risk estimates derived from sex-specific multiple logistic functions estimated by the authors using data from the Framingham Heart Study or generated by the Centers for Disease Control-sponsored Risk Factor Update Project (RFUP).[10] A random sample of 240 observations from the Framingham data set serve as test cases to correlate risk estimates from the criterion logistic models with those from each of 41 existing HRAs.

This pragmatic approach yields some very useful information. First, many of the HRA procedures are more highly correlated with the criterion logistic models than the criterion models are with each other, demonstrating that estimates from many HRAs agree rather well with estimates obtained from conventional epidemiologic approaches. Evidently both the HRA risk estimation algorithms and the data bases they use produce reasonable results. This finding is welcome from a quality-assurance perspective. Secondly, and somewhat surprising, Norman Gesner's "credit-debit method" of combining individual risk factors,[3] which although intuitively appealing has been widely criticized, is apparently capable of yielding risk estimates similar to those from logistic models.

The study's limitations are also important. First, the Framingham data are a major source both for the criterion models and for the CHD risk factor data in many HRAs, so that to some extent the correlations tell us how much has been lost in translation rather than how valid are estimates in reference to an independent criterion.[15] On the other hand, the Framingham data are very consistent with those from other major US cohort studies.[14]

Second, we do not know the effect on the correlations of the omission of risk characteristics used in many HRAs but for which data were not available for the test cases (e.g., exercise, family history of heart disease). It is possible that setting these to average levels, although a reasonable approach, inflates the correlations.

Third, since the test cases were randomly selected, the correlations between HRAs and the criterion models tell us only about overall performance. Conceivably there could still be substantial discrepancies for individuals at high- or low-risk levels, or with particular risk indicator patterns. Chaves[13] generated hypothetical test cases by randomly varying risk indicator levels, and observed the influence of gender, blood pressure, diabetes, smoking, and cholesterol to vary substantially across instruments, with marked differences in risk estimates for hypothetical individuals.

Also, as the authors clearly indicate, the findings apply only to CHD, HRA's (and epidemiology's) strongest suit in terms of knowledge of risk characteristics and availability of datasets. For other major causes of death estimated in HRA, quantitative data on risk characteristics are often severely limited[19]; validity in estimating overall mortality risk should therefore be lower. Chaves[13] observed different rankings of the top five causes of death across different HRA instruments, apparently due to the use of mortality statistics from different time periods.

More troubling is the potential for overinterpretation of the findings of this study, particularly since the term "validity" so strongly connotes desirability. Based on their measurement validity perspective, the authors identify several HRA characteristics as reducing the "validity" of HRA risk scores: 1) use of an additive weighting method to generate an arbitrary risk scale; 2) having a limited range of risk estimates, due to including fewer disease determinants measured in broad categories; 3) not taking age into consideration in the HRA estimation procedure. There are a number of reasons why these characteristics do not necessarily imply deficiency, although they certainly do reduce precision and correlations to more sophisticated risk estimation models.

Many of the HRA instruments that had low correlations generated "general health scores" or other arbitrary risk scales. It is only reasonable to expect these instruments to be less strongly correlated with the criterion logistic models than HRAs designed to estimate CHD risk. But even where heart disease risk is the focus, a degree of inaccuracy may be completely acceptable if the instrument does not purport to be particularly accurate or precise. Indeed, a lesser degree of inaccuracy in an instrument with greater appearances of precision may be more misleading.[20]

Also, the impact of not taking age into consideration is clear in respect to risk estimates but not in respect to HRA validation or application. Since inclusion of age, as the strongest correlate of CHD risk, greatly strengthens precision of estimation, it could be argued that the correlations presented by Smith, *et al,*[11] underplay the effects of the modifiable risk characteristics which are presumably the most relevant aspects of HRA. In the Chaves, *et al,* study,[13] the contribution of age largely explained the almost identical rank-ordering of heart disease risk across instruments. More important, inclusion of age and the generation of appropriately skewed, rather than normally distributed, risk scores may or may not improve the utility of HRA feedback in actual use. For these reasons, the term "validity" should be read in its narrowest sense.

HRA, as a vehicle for what might be termed "prospective health assessment," potentially has a number of very desirable qualities for clinicians and health educators: preventive orientation, systematic approach, ability to emphasize modifiable factors, and grounding in current scientific knowledge. A recent conference highlighted the diversity of settings and uses of HRA,** each involving different objec-

**Personal Health Risk Assessment Methods in Health Hazard/Health Risk Appraisal: A Research Agenda. Wayzetta, Minnesota, September 7–9, 1986; conducted by the Association for Health Services Research and soon to be reported in *Health Services Research.*

tives, raising different concerns for validity, and entailing different requirements for effective presentation of results. A major question arising at the conference is what is the value of, or when is it valuable to have, quantitative estimates of absolute risk—as opposed to relative risk, risk scores, health scores, and other less quantitatively ambitious measures—given the limitations in scientific knowledge and risk estimation methodology.[21] Although general accuracy in respect to risk, risk characteristics, relative risk, and behavioral recommendations is certainly essential, sophistication and precision in risk estimation are not necessarily the measure of quality of HRA, however valuable they may be for other purposes. The Smith, *et al*, study[11] is a most helpful contribution to assessing HRA's basic accuracy.

REFERENCES

1. Beery WL, Schoenbach VJ, Wagner EH, and associates: Health Risk Appraisal: Methods and Programs, with Annotated Bibliography. DHHS Pub. No. (PHS)86-3396. Hyattsville, MD: National Center for Health Services Research and Health Care Technology, 1986.
2. Hall JH, Zwemer JD: Prospective Medicine. Indianapolis: Methodist Hospital of Indiana, 1979.
3. Gesner NB: Derivation of risk factors from comparative data. In: Proceedings of the Seventh Annual Meeting of the Society of Prospective Medicine. Indianapolis: Methodist Hospital of Indiana, 1971; 48.
4. Robbins LC, Hall JH: How to Practice Prospective Medicine. Indianapolis: Methodist Hospital of Indiana, 1970.
5. Lalonde M: A New Perspective on the Health of Canadians: A Working Document. Ottawa: Information Canada, 1975.
6. Healthfinder. National Health Information Clearinghouse, Office of Disease Prevention and Health Promotion, US Department of Health and Human Services, Washington, DC, 1985.
7. National Worksite Health Promotion Survey: A summary of results. Office of Disease Prevention and Health Promotion, Department of Health and Human Services, Washington, DC (Monograph in press).
8. Brown KS, Nabert W: Evaluation of the existing method for calculating health hazard appraisal age. Report for Non-Medical Use of Drugs Directorate, Health Promotion Branch, Health and Welfare, Canada, by the University of Waterloo, Ontario, August 31, 1977.
9. Spasoff RA, McDowell IW, Wright PA, Dunkley GC: Proposals for a Revised Evalu Life. Final report submitted to Health and Welfare, Canada, 1981.
10. Breslow L, Fielding J, Afifi AA, *et al:* Risk Factor Update Project: Final report. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control, Center for Health Promotion and Education, 1985 (Contract No. 200-80-0527).
11. Smith KW, McKinlay SM, Thorington BD: The validity of health risk appraisal instruments for assessing coronary heart disease risk. Am J Public Health 1987; 77:419–424.
12. Wiley JA: Predictive risk factors do predict life events. In: Proceedings of the 16th Annual Meeting of the Society of Prospective Medicine. Bethesda, MD: Society of Prospective Medicine, 1981; 75–79.
13. Chaves MA, Jennings SE, McKinlay SM, McKinlay JB: Cardiovascular risk: differences among health hazard appraisals. In: Proceedings of the 20th Annual Meeting of the Society of Prospective Medicine. Washington, DC: Society of Prospective Medicine, 1985; 25–27.
14. Kannel WB, McGee DL: Composite scoring: methods and predictive validity. Insights from the Framingham Study. Health Serv Res (in press).
15. McDowell I: Re: On the validity of health risk appraisal. (Letter) Nurs Res 1982; 31:37.
16. McDowell I: Dilemmas in evaluating behaviour-change strategies in disease prevention. (Letter) J Chronic Dis 1984; 37:144–145.
17. Gordon T, Kannel WB, Halperin M: Predictability of coronary heart disease. J Chronic Dis 1979; 32:427–440.
18. Nunnally JC: Psychometric Theory. New York: McGraw-Hill, 1967.
19. Schoenbach VJ, Wagner EH, Karon JM: The use of epidemiologic data for personal risk assessment in health hazard/health risk appraisal programs. J Chronic Dis 1983; 36:625–638.
20. Safer MA: An evaluation of the health hazard appraisal based on survey data from a randomly selected population. Public Health Rep 1982; 97:31–37.
21. Schoenbach VJ, Wagner EH, Beery WL: Health risk appraisal: review of evidence for effectiveness. Health Serv Res (In press).

VICTOR J. SCHOENBACH, PHD

Address reprint requests to Victor J. Schoenbach, PhD, Assistant Professor, Department of Epidemiology, School of Public Health, University of North Carolina, Rosenau 201H, Chapel Hill, NC 27514-6201. Dr. Schoenbach is also Research Associate, Health Services Research Center, University of North Carolina at Chapel Hill.

New Health Risk Appraisal Nearly Completed

A completely revised Health Risk Appraisal (HRA) is nearing completion at the Carter Center of Emory University with technical sponsorship by the American Public Health Association, the Centers for Disease Control, and several other health agencies. The new HRA is designed to address many of the points made by Dr. Schoenbach in the above editorial. The objective is a fully documented computer program for the public domain, with open architecture to accommodate improved data as the technology of prevention matures. The completed software will be released at the annual meeting of the Society of Prospective Medicine to be held in Atlanta, September 17–20, 1987.

For further information about the HRA, contact: Marjorie Bland, Coordinator, Health Risk Appraisal Project, Carter Center, Emory University, 1989 North Williamsburg Drive, Suite E, Decatur, GA 30033. Telephone 404/321–4104.